(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 November 2003 (27.11.2003)

PCT

(10) International Publication Number
WO 03/098186 A2

(51) International Patent Classification[7]: G01N

(21) International Application Number: PCT/US03/15269

(22) International Filing Date: 16 May 2003 (16.05.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/150,559 17 May 2002 (17.05.2002) US

(71) Applicant *(for all designated States except US)*: **THE REGENTS OF THE UNIVERSTIY OF CALIFORNIA** [US/US]; 1111 Franklin Street, Fifth Floor, Oakland, CA 94607-5200 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **WILKINS, Thea, A.** [US/US]; 262 Palm Avenue, Woodland, CA 95695 (US).

(74) Agents: **HAJOS, Brigitte, A.** et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, 8th Floor, San Francisco, CA 94111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/098186 A2

(54) Title: BIOENGINEERING COTTON FIBER PROPERTIES

(57) Abstract: The present invention provides plant fiber expansion (FE) genes that encode FE polypeptides, such as phosphoenol pyruvate carboxylase (PEPcase), expansin, endoglucanase, xyloglucan endoglycosyltransferase (XET), and pectin methyl esterase (PME). The invention further provides fiber-specific promoters. Still further, the invention provides molecular strategies for modulating fiber quality and yield in fiber producing plants by modulating expression of FE genes or mutant forms of FE genes.

BIOENGINEERING COTTON FIBER PROPERTIES

RELATED APPLICATION INFORMATION

[0001] This application is a continuation-in-part of U.S. Application Serial No. 10/150,559 filed May 17, 2002. This application is incorporated by reference herein.

FIELD OF THE INVENTION

[0002] The present invention is directed to plant genetic engineering. In particular, it relates to the isolation of nucleic acid molecules that modulate fiber quality and yield, and the use of these nucleic acid molecules to produce transgenic plants with varied cotton fiber characteristics and quality.

BACKGROUND OF THE INVENTION

[0003] Cotton is a widely used textile fiber. For example, cotton textiles are used for clothing, home furnishings, blanket fills, toiletry products, industrial garments, etc. The expansive utility of cotton textile products is attributed to the relative ease of cotton production compared to other fibers and their appealing properties. As clothing, cotton fabrics are comfortable to wear because they are soft and breathable. Furthermore, cotton fibers are highly absorptive and possess good wicking properties, thereby allowing the use of the fibers in absorbent articles.

[0004] Although cotton is one of the most popular textile fibers used, it has many disadvantages. For example, cotton fabrics become worn out readily after several cycles of laundering. This is because, cotton fibers break or pill due to mechanical agitation during wash and form a lint on the surface of the fabric. In another example, cotton fibers tend to shrink significantly compared to synthetic fibers, even after several cycles of laundry. The shrinkage of cotton textile products, in particular clothing, poses a dilemma for consumers, because the consumers can not readily determine how much their newly purchased cotton clothing will shrink and if the clothing will fit on them to their satisfaction after a few cycles of wash. In yet another example, cotton fabrics tend to wrinkle easily, and require a great deal of care to maintain their shape.

[0005] In order to overcome these disadvantages, manufacturers often pre-treat cotton fibers and fabrics. For example, to control lint formation, cotton seeds are delinted prior to a brush delinter, or cotton fabrics are treated with a cellulase solution to remove lint precursors. To reduce wrinkle formation, manufacturers treat cotton fabrics with crosslinking agents, such as formaldehyde. However, these additional processes to treat cotton fibers or fabrics add cost to the manufacture of cotton textile products. Furthermore, chemicals added during the manufacture of cotton fabrics and fibers tend to wash out during laundering and lose their effect over time.

[0006] Thus, there is a need to improve the quality of cotton textile products. It would be desirable to avoid using any additives in improving the quality of cotton textile products, because they lose their effect over time, especially after repetitive laundering. Chemical additives may also be toxic to human body. It would also be desirable to reduce any additional processing steps so that the manufacture of cotton textile products will be cost effective. One way to resolve these problems is by improving the quality of cotton fibers themselves, so that the need for additional processing steps is eliminated. Thus, there is a need to improve the cotton fiber characteristics, such as fiber strength, fiber length and fineness.

SUMMARY OF THE INVENTION

[0007] The present invention provides isolated nucleic acid molecules comprising a FE polynucleotide sequences. Examples of nucleic acids of the invention include phosphoenol pyruvate carboxylase (PEPcase) sequences at least about 60% identical to SEQ ID NO:1, expansin sequences at least about 60% identical to SEQ ID NO:3, endoglucanase sequences at least about 60% identical to SEQ ID NO: 5, xyloglucan endoglycosyltransferse (XET) sequences at least about 60% identical to SEQ ID NO: 7, and pectin methyl esterase (PME) sequences at least about 60% identical to SEQ ID NO: 9. The isolated nucleic acid molecules of the invention may further comprise a plant promoter operably linked to the FE polynucleotide. The promoter may be, for example, a tissue-specific promoter, in particular, a fiber-specific promoter. The FE polynucleotides may be linked to the promoter in a sense or an antisense orientation.

[0008] The invention also provides transgenic plants comprising an expression cassette containing a plant promoter operably linked to a heterologous FE polynucleotide sequence of the invention.

[0009] The invention further provides methods of modulating fiber quality in a plant. The methods comprise introducing into the plant an expression cassette containing a plant promoter operably linked to a heterologous FE polynucleotide sequence of the invention. The plant may be any plant and is usually a member of the genus *Gossypium*. In the methods the expression cassette can be introduced into the plant through a sexual cross or using genetic engineering techniques.

[0010] The invention also encompasses a method of modulating fiber yield and/or fiber quality in a plant. The method comprises introducing into a plant an expression cassette containing a plant promoter operably linked to a heterologous expansin polynucleotide sequence, which is at least about 80% identical to SEQ ID NO: 3. The heterologous expansin polynucleotide encodes an expansin polypeptide. The heterologous expansin polypeptide may include an amino acid sequence of SEQ ID NO: 4. The heterologous expansin polynucleotide may include a nucleic acid sequence of SEQ ID NO: 3. The method may further contemplate the use of a plant promoter that is an enhanced cotton fiber specific promoter. The plant may be any plant and is usually a member of the genus *Gossypium*.

DEFINITIONS

[0011] The phrase "nucleic acid sequence" refers to a single or double-stranded polymer of deoxyribonucleotide or ribonucleotide bases read from the 5' to the 3' end. It includes chromosomal DNA, self-replicating plasmids, infectious polymers of DNA or RNA and DNA or RNA that performs a primarily structural role.

[0012] A "promoter" is defined as an array of nucleic acid control sequences that direct transcription of an operably linked nucleic acid. As used herein, a "plant promoter" is a promoter that functions in plants, even though obtained from other organisms, such as plant viruses. Promoters include necessary nucleic acid sequences near the start site of transcription, such as, in the case of a polymerase II type promoter, a TATA element. A promoter also optionally includes distal enhancer or repressor elements, which can be located as much as several thousand base pairs from the start site of transcription. The

term "operably linked" refers to a functional linkage between a nucleic acid expression control sequence (such as a promoter, or array of transcription factor binding sites) and a second nucleic acid sequence, wherein the expression control sequence directs transcription of the nucleic acid corresponding to the second sequence.

[0013] The term "plant" includes whole plants, plant organs (*e.g.*, leaves, stems, flowers, roots, etc.), seeds and plant cells and progeny of same. The class of plants that can be used in the method of the invention is generally as broad as the class of higher plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), as well as gymnosperms. It includes plants of a variety of ploidy levels, including polyploid, diploid, haploid and hemizygous.

[0014] A polynucleotide sequence is "heterologous to" an organism or a second polynucleotide sequence if it originates from a foreign species, or, if from the same species, is modified from its original form. For example, a promoter operably linked to a heterologous coding sequence refers to a coding sequence from a species different from that from which the promoter was derived, or, if from the same species, a coding sequence which is different from any naturally occurring allelic variants.

[0015] A polynucleotide "exogenous to" an individual plant is a polynucleotide which is introduced into the plant by any means other than by a sexual cross. Examples of means by which this can be accomplished are described below, and include *Agrobacterium*-mediated transformation, biolistic methods, electroporation, and the like. Such a plant containing the exogenous nucleic acid is referred to here as an R_1 generation transgenic plant. Transgenic plants that arise from sexual cross or by selfing are descendants of such a plant.

[0016] "FE" is an acronym for fiber expansion, and the term is used generically to refer to properties of cotton fibers controlled by the polynucleotides and polypeptides of the present invention. For example, an FE polynucleotide refers to nucleic acids encoding FE polypeptides, such as phosphoenol pyruvate carboxylase (PEPcase), expansin, endoglucanase, xyloglucan endoglycosyltransferse (XET), and pectin methyl esterase (PME).

[0017] "Phosphoenol pyruvate carboxylase" or "PEPcase" refers to an enzyme that regulates synthesis of malate. Malate is a primary osmoregulatory solute involved in maintaining cell turgor during fiber expansion. Thus, a "phosphoenol pyruvate

carboxylase polynucleotide" or "PEPcase polynucleotide" of the invention is a subsequence or full length polynucleotide sequence of a gene which, when present in a transgenic plant, can be used to modify fiber quality (*e.g.*, fiber length, fiber strength, or fiber fineness) and which is at least about 60%, 70%, 80%, 85%, 90%, 95%, 98% or more identical to SEQ ID NO:1. A PEPcase polynucleotide typically comprises or consists of a coding region of at least about 30-40 nucleotides to about 3400 nucleotides in length. Usually, the nucleic acids are from about 100 to about 500 nucleotides, often from about 500 to about 1500 nucleotides in length or from about 1500 nucleotides in length to about 3400 nucleotides in length.

[0018] "Expansin" refers to an enzyme that influences cross-linking relationships in the cell wall and allow cell wall components to "slip" during fiber expansion, thereby allowing the fibers to increase in length. Thus, an "expansin polynucleotide" of the invention is a subsequence or full length polynucleotide sequence of a gene which, when present in a transgenic plant, can be used to modify fiber quality (*e.g.*, fiber length, fiber strength, or fiber fineness) and fiber yield (fiber weight per seed); and which is at least about 60%, 70%, 80%, 85%, 90%, 95%, 98% or more identical to SEQ ID NO:3. An expansin polynucleotide typically comprises or consists of a coding region of at least about 30-40 nucleotides to about 1154 nucleotides in length. Usually, the nucleic acids are from about 100 to about 500 nucleotides, often from about 500 to about 1154 nucleotides in length.

[0019] "Endoglucanase" refers to a type of cellulase that cleaves glucan cellulose, thereby controlling the length of cellulose polymers. Thus, an "endoglucanase polynucleotide" of the invention is a subsequence or full length polynucleotide sequence of a gene which, when present in a transgenic plant, can be used to modify fiber quality (*e.g.*, fiber length, fiber strength, or fiber fineness) and which is at least about 60%, 70%, 80%, 85%, 90%, 95%, 98% or more identical to SEQ ID NO:5. An endoglucanase polynucleotide typically comprises or consists of a coding region of at least about 30-40 nucleotides to about 2386 nucleotides in length. Usually, the nucleic acids are from about 100 to about 500 nucleotides, often from about 500 to abou 1500 nucleotides in length or from about 1500 nucleotides in length to about 2386 nucleotides in length..

[0020] "Xyloglucan endoglycosyltranferase" or "XET" refers to an enzyme that modifies cross-linking relationships between cellulose microfibrils and the xyloglucan

matrix, and loosens the cell wall. Thus, a "xyloglucan endoglycosyltransferase" or "XET" of the invention is a subsequence or full length polynucleotide sequence of a gene which, when present in a transgenic plant, can be used to modify fiber quality (*e.g.*, fiber length, fiber strength, or fiber fineness) and which is at least about 60%, 70%, 80%, 85%, 90%, 95%, 98% or more identical to SEQ ID NO:7. A XET polynucleotide typically comprises or consists of a coding region of at least about 30-40 nucleotides to about 1179 nucleotides in length. Usually, the nucleic acids are from about 100 to about 500 nucleotides, often from about 500 to about 1179 nucleotides in length.

[0021] "Pectin methyl esterase" or "PME" refers to an enzyme that is involved in esterification of the pectin matrix. Thus, a "pectin methyl esterase" or "PME" of the invention is a subsequence or full length polynucleotide sequence of a gene which, when present in a transgenic plant, can be used to modify fiber quality (*e.g.*, fiber length, fiber strength, or fiber fineness) and which is at least about 60%, 70%, 80%, 85%, 90%, 95%, 98% or more identical to SEQ ID NO:9. A PME polynucleotide typically comprises or consists of a coding region of at least about 30-40 nucleotides to about 1702 nucleotides in length. Usually, the nucleic acids are from about 100 to about 500 nucleotides, often from about 500 to about 1702 nucleotides in length.

[0022] For any polypeptides described above, one of skill in the art will recognize that in light of the present disclosure, various modifications (*e.g.*, substitutions, additions, and deletions) can be made to the polypeptide sequences without substantially affecting their function. These variations are within the scope of the present invention.

[0023] In the case of both expression of transgenes and inhibition of endogenous genes (*e.g.*, by antisense, or sense suppression) one of skill will recognize that the inserted polynucleotide sequence need not be "identical," but may be only "substantially identical" to a sequence of the gene from which it was derived.

[0024] The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection.

[0025] The phrase "substantially identical," in the context of two nucleic acids or polypeptides, refers to two or more sequences or subsequences that have at least about

60%, or at least about 70%, preferably at least about 80%, most preferably at least about 90-98% nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using one of the following sequence comparison algorithms or by visual inspection. Preferably, the substantial identity exists over a region of the sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably the sequences are substantially identical over at least about 150 residues. In a most preferred embodiment, the sequences are substantially identical over the entire length of the coding regions.

[0026] For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0027] Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (*see generally, Current Protocols in Molecular Biology*, F.M. Ausubel *et al.*, eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (1995 Supplement) (Ausubel)).

[0028] Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul *et al.* (1990) *J. Mol. Biol.* 215: 403-410 and Altschuel *et al.* (1977) *Nucleic Acids Res.* 25: 3389-3402, respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence,

which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (*see* Henikoff & Henikoff, *Proc. Natl. Acad. Sci. USA* 89:10915 (1989)).

[0029] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.*, Karlin & Altschul, *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

[0030] A further indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another

indication that two nucleic acid sequences are substantially identical is that the two molecules hybridize to each other under stringent conditions, as described below.

[0031] "Conservatively modified variants" applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refers to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence herein which encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

[0032] As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alters, adds or deletes a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art.

[0033] The following six groups each contain amino acids that are conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);
2) Aspartic acid (D), Glutamic acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and

6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).
(*see, e.g.*, Creighton, *Proteins* (1984)).

[0034] An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions, as described below.

[0035] The phrase "selectively (or specifically) hybridizes to" refers to the binding, duplexing, or hybridizing of a molecule only to a particular nucleotide sequence under stringent hybridization conditions when that sequence is present in a complex mixture (*e.g.*, total cellular or library DNA or RNA).

[0036] The phrase "stringent hybridization conditions" refers to conditions under which a probe will hybridize to its target subsequence, typically in a complex mixture of nucleic acid, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures. An extensive guide to the hybridization of nucleic acids is found in Tijssen, *Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Probes*, "Overview of principles of hybridization and the strategy of nucleic acid assays" (1993). Generally, highly stringent conditions are selected to be about 5-10°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength pH. Low stringency conditions are generally selected to be about 15-30 °C below the T_m. The T_m is the temperature (under defined ionic strength, pH, and nucleic concentration) at which 50% of the probes complementary to the target hybridize to the target sequence at equilibrium (as the target sequences are present in excess, at T_m, 50% of the probes are occupied at equilibrium). Stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion concentration (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30°C for short probes (*e.g.*, 10 to 50 nucleotides) and at least about 60°C for long probes (*e.g.*, greater than 50 nucleotides). Stringent conditions may also be achieved with the

addition of destabilizing agents such as formamide. For selective or specific hybridization, a positive signal is at least two times background, preferably 10 time background hybridization.

[0037] Nucleic acids that do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, for example, when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code. In such cased, the nucleic acids typically hybridize under moderately stringent hybridization conditions.

[0038] In the present invention, genomic DNA or cDNA comprising nucleic acids of the invention can be identified in standard Southern blots under stringent conditions using the nucleic acid sequences disclosed here. For the purposes of this disclosure, suitable stringent conditions for such hybridizations are those which include a hybridization in a buffer of 40% formamide, 1 M NaCl, 1% SDS at 37°C, and at least one wash in 0.2X SSC at a temperature of at least about 50°C, usually about 55°C to about 60°C, for 20 minutes, or equivalent conditions. A positive hybridization is at least twice background. Those of ordinary skill will readily recognize that alternative hybridization and wash conditions can be utilized to provide conditions of similar stringency.

[0039] A further indication that two polynucleotides are substantially identical is if the reference sequence, amplified by a pair of oligonucleotide primers, can then be used as a probe under stringent hybridization conditions to isolate the test sequence from a cDNA or genomic library, or to identify the test sequence in, *e.g.*, a northern or Southern blot.

[0040] "Fiber specific" promoter refers to promoters that preferentially promote gene expression in fiber cells over other cell types.

DETAILED DESCRIPTION

[0041] This invention provides plant FE genes that encode FE polypeptides, such as phosphoenol pyruvate carboxylase (PEPcase), expansin, endoglucanase, xyloglucan endoglycosyltransferase (XET), and pectin methyl esterase (PME). The invention further provides fiber-specific promoters. Still further, the invention provides molecular strategies for modulating fiber quality and yield in fiber producing plants by modulating expression of FE genes or mutant forms of FE genes.

[0042] Important fiber properties, such as fiber length, strength, and fineness, are determined by rate and duration of fiber expansion. Fiber expansion is, in turn, dependent primarily on cell turgor, the driving force of fiber expansion, and the extensibility of the cell wall. By manipulating genes that regulate these critical processes, fiber growth and fiber properties can be modified.

[0043] There are several genes encoding enzymes that are involved in maintaining turgor during fiber expansion. One such enzyme is phosphoenol pyruvate carboxylase (PEPcase). A PEPcase regulates synthesis of malate, which is a primary osmoregulatory solute involved in maintaining cell turgor during fiber expansion. By modulating the expression of PEPcase, the rate and/or duration of fiber expansion and fiber length can be regulated.

[0044] There are also several enzymes that regulate extensibility of fiber cell walls. These include: 1) expansins; 2) endoglucanases; 3) xyloglucan endoglycosyltransferases (XET); and 4) pectin methyl esterases (PME).

[0045] Expansins are thought to function in plant cell growth, cell wall disassembly, and cell separation. In particular, expansins influence cross-linking relationships in the cell wall and allow cell wall components to "slip" during fiber expansion, thereby allowing the fibers to increase in length. There are two classes of expansins that are currently recognized, alpha-expansins (EXP) and beta-expansins (EXPB). Alpha-expansins are a conserved group of proteins that function in cell wall enlargement and possibly other developmental processes including cell wall disassembly and cell separation. Beta-expansins are known to be secreted by grass pollen and have cell wall loosening effects on grass cells walls. It appears that one role of beta-expansins is to soften the stigma and stylar tissues to speed penetration of the pollen tubes through the maternal tissues to the ovule. Other beta-expansins are found in young grass seedlings and non-pollen tissue, thus, hinting at a variety of developmental functions of these proteins.

[0046] Generally, expansins are proteins with relatively conserved motifs. The mature protein may include two domains, a cysteine-rich region with limited sequence similarity to family-45 endoglucanases (EG45-like domain) and a tryptophan-rich carboxy terminus that may function as a putative polysaccharide-binding domain. Alpha- and beta-expansins share only about 20 to 25% amino acid sequence identity. The regions of

identity are found throughout the protein backbone and include 6 of the 8 conserved cysteine residues in the cysteine-rich region, and the tryptophans near the carboxy terminus. Beta-expansins are glycosylated proteins while alpha-expansins are not glycosylated.

[0047] Studies in the small flowering plant, *Arabidopsis thaliana (A. thaliana)*, have shown that the expansin gene family is extensive. *A. thaliana* contains 26 genes that encode alpha-expansin proteins, wherein these proteins are designated as EXP1 through EXP26. Most of the alpha-expansin genes have a conserved intron structure (*i.e.*, two introns), wherein the intron lengths range from about 90 bp to about 500 bp. Some alpha-expansin genes (*e.g.*, EXP17 through EXP26) are missing intron 2; EXP10 has an additional intron in the 5' untranslated region. *A. thaliana* also contains 5 genes that encode beta-expansin proteins which are designated EXPB1 through EXPB5. These genes typically contain three introns, of which two introns are conserved in alpha expansins, and a third intron that is not found in alpha-expansins and may be located before or after intron 2.

[0048] Alpha- and beta-expansin genes have been investigated in a number of species, including cucumber, rice, pea, tomato, strawberry, apricot, pine, tobacco, fern, and maize. Expansin proteins are also found in cotton such as in *Gossypium hirsutum* (*G. hirsutum*), in fact, there are two major isoforms such as GhEXP1 and GhEXP2. GhEXP1 (EpGhEX1), a full length cDNA, encodes a 258 amino acid alpha-expansin protein with a N-terminal signal peptide. The corresponding transcript is abundant in cotton fiber cells but absent in all other tissues tested. This gene is presumed to be developmentally regulated during fiber elongation in *G. hirsutum* (see Orford *et al.* (1998) *Biochem. Biophys. Acta.* 1398(3):342-346).

[0049] The GhEXP2 cDNA (SEQ ID NO: 3) encodes a novel alpha expansin protein (SEQ ID NO: 4) from *G. hirsutum*. This novel GhEXP2 differs structurally from GhEXP1 and plays an important role in cell elongation. Overexpression of GhEXP2 shows significant gains in cotton yield and fiber quality (see Example 2, *vide infra*).

[0050] Other enzymes are involved in cell wall relaxation during fiber expansion. For example, an endoglucanase is a cellulase that cleaves glucan cellulose, thereby controlling the length of cellulose polymers. Changing the cellulose polymer length in primary cell walls of developing fibers can strongly influence fiber length. In another

example, XETs are important in cell wall loosening, by changing cross-linking relationships between cellulose microfibrils and the xyloglucan matrix. In yet another example, PMEs are enzymes that are involved in esterification of the pectin matrix. The pectin matrix is highly esterified during rapid fiber expansion. When esterified pectin fraction is deesterified, it results in increased cell wall rigidity during the termination of fiber expansion. Not wishing to be bound by a theory, delaying the deesterification of this pectin fraction can increase the duration of fiber expansion, and hence, fiber length.

[0051] A single FE or any combinations of the FE nucleic acids encoding the above enzymes can be introduced into a plant to modulate the quality of fibers. Preferably, a fiber-specific promoter is used to express the FE nucleic acids only in fibers of plants. More preferably, an inducible fiber specific promoter is used to express these genes during appropriate developmental stages most likely to result in increased fiber growth.

ISOLATION OF NUCLEIC ACIDS

[0052] Generally, the nomenclature and the laboratory procedures in recombinant DNA technology described below are those well known and commonly employed in the art. Standard techniques are used for cloning, DNA and RNA isolation, amplification and purification. Generally enzymatic reactions involving DNA ligase, DNA polymerase, restriction endonucleases and the like are performed according to the manufacturer's specifications. These techniques and various other techniques are generally performed according to Sambrook *et al.*, *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, (1989) or Current Protocols in Molecular Biology Volumes 1-3, John Wiley & Sons, Inc. (1994-1998).

[0053] The isolation of nucleic acids may be accomplished by a number of techniques. For instance, oligonucleotide probes based on the sequences disclosed here can be used to identify the desired gene in a cDNA or genomic DNA library. To construct genomic libraries, large segments of genomic DNA are generated by random fragmentation, *e.g.* using restriction endonucleases, and are ligated with vector DNA to form concatemers that can be packaged into the appropriate vector. To prepare a cDNA library, mRNA is isolated from the desired organ, such as leaves, and a cDNA library which contains gene transcripts is prepared from the mRNA. Alternatively, cDNA may be prepared from mRNA extracted from other tissues in which genes of interest or their homologs are expressed.

[0054] The cDNA or genomic library can then be screened using a probe based upon the sequence of a cloned gene disclosed here. Probes may be used to hybridize with genomic DNA or cDNA sequences to isolate homologous genes in the same or different plant species. Alternatively, antibodies raised against a polypeptide of interest can be used to screen an mRNA expression library.

[0055] Alternatively, the nucleic acids of interest can be amplified from nucleic acid samples using amplification techniques. For instance, polymerase chain reaction (PCR) technology can be used to amplify the sequences of genes directly from genomic DNA, from cDNA, from genomic libraries or cDNA libraries. PCR and other *in vitro* amplification methods may also be useful, for example, to clone nucleic acid sequences that code for proteins to be expressed, to make nucleic acids to use as probes for detecting the presence of the desired mRNA in samples, for nucleic acid sequencing, or for other purposes. For a general overview of PCR, *see PCR Protocols: A Guide to Methods and Applications*. (Innis, M, Gelfand, D., Sninsky, J. and White, T., eds.), *Academic Press*, San Diego (1990). Appropriate primers and probes for identifying sequences from plant tissues are generated from comparisons of the sequences provided herein (*e.g.* SEQ ID NO: 1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, etc.).

[0056] Polynucleotides may also be synthesized by well-known techniques, as described in the technical literature. *See, e.g.*, Carruthers *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* 47:411-418 (1982), and Adams *et al.*, *J. Am. Chem. Soc.* 105:661 (1983). Double stranded DNA fragments may then be obtained either by synthesizing the complementary strand and annealing the strands together under appropriate conditions, or by adding the complementary strand using DNA polymerase with an appropriate primer sequence.

INCREASING LEVELS OF GENE EXPRESSION IN PLANT FIBERS

[0057] The isolated nucleic acid sequences prepared as described herein can be used in a number of techniques. For example, the isolated nucleic acids can be introduced into plants to enhance endogenous gene expression. A particularly useful gene for this purpose is the FE genes shown in SEQ ID NO: 1, 3, 5, 7, and 9. In one embodiment, more than one gene can be introduced into plants. For example, expansins and endoglucanases can be expressed in plant fibers, thereby modifying crosslinking relationships and the cellulose polymer length in primary cell walls. Preferably, fiber

tissues are targeted to increase expression FE genes. Fibers can be targeted at all times during the life of the plant *e.g.*, using a constitutive promoter, or transiently, *e.g.*, using a transiently active or an inducible promoter.

[0058] Isolated nucleic acids prepared as described herein can be used to introduce expression of particular FE nucleic acids to enhance endogenous gene expression. Enhanced expression will lead to increased fiber quality, such as fiber length, strength, and fineness; and increased fiber yield. Thus, plants comprising these constructs are particularly useful for producing fibers with improved properties for textile products. Where overexpression of a gene is desired, the desired gene from a different species may be used to decrease potential sense suppression effects. One of skill will recognize that the polypeptides encoded by the genes of the invention, like other proteins, have different domains which perform different functions. Thus, the gene sequences need not be full length, as long as the desired functional domain of the protein is expressed.

[0059] Modified protein chains can also be readily designed utilizing various recombinant DNA techniques well known to those skilled in the art and described in detail below. For example, the chains can vary from the naturally occurring sequence at the primary structure level by amino acid substitutions, additions, deletions, and the like. These modifications can be used in a number of combinations to produce the final modified protein chain.

[0060] In another embodiment, modified forms of genes disclosed here can be used that have increased activity *in vivo*. For example, endoglucanase mutants that elongate the cellulose polymer length can be created and used to produce transgenic plants. Additional hyperactive forms can be readily identified, *e.g.*, by screening for modified forms of FE enzymes with an increased ability to modify fiber quality such as fiber length, strength, and fineness.

[0061] In another embodiment, endogenous gene expression can be targeted for modification. Methods for introducing genetic mutations into plant genes and selecting plants with desired traits are well known. For instance, seeds or other plant material can be treated with a mutagenic chemical substance, according to standard techniques. Such chemical substances include, but are not limited to, the following: diethyl sulfate, ethylene imine, ethyl methanesulfonate and N-nitroso-N-ethylurea. Alternatively, ionizing radiation from sources such as X-rays or gamma rays can be used.

[0062] Alternatively, homologous recombination can be used to induce targeted gene modifications by specifically targeting the FE gene *in vivo* (*see, generally*, Grewal and Klar, *Genetics* 146: 1221-1238 (1997) and Xu *et al., Genes Dev.* 10: 2411-2422 (1996)). Homologous recombination has been demonstrated in plants (Puchta *et al., Experientia* 50: 277-284 (1994), Swoboda *et al., EMBO J.* 13: 484-489 (1994); Offringa *et al., Proc. Natl. Acad. Sci. USA* 90: 7346-7350 (1993); and Kempin *et al., Nature* 389:802-803 (1997)).

[0063] In applying homologous recombination technology to the genes of the invention, mutations in selected portions of a FE gene sequence (including 5' upstream, 3' downstream, and intragenic regions) such as those disclosed herein are made *in vitro* and then introduced into the desired plant using standard techniques. Since the efficiency of homologous recombination is known to be dependent on the vectors used, use of dicistronic gene targeting vectors as described by Mountford *et al., Proc. Natl. Acad. Sci. USA* 91: 4303-4307 (1994); and Vaulont *et al., Transgenic Res.* 4: 247-255 (1995) are conveniently used to increase the efficiency of selecting for altered FE expression in transgenic plants. The mutated gene will interact with the target wild-type gene in such a way that homologous recombination and targeted replacement of the wild-type gene will occur in transgenic plant cells, resulting in increased FE activity.

[0064] Alternatively, oligonucleotides composed of a contiguous stretch of RNA and DNA residues in a duplex conformation with double hairpin caps on the ends can be used. The RNA/DNA sequence is designed to align with the sequence of the target gene and to contain the desired nucleotide change. Introduction of the chimeric oligonucleotide on an extrachromosomal T-DNA plasmid results in efficient and specific FE gene conversion directed by chimeric molecules in a small number of transformed plant cells. This method is described in Cole-Strauss *et al., Science* 273:1386-1389 (1996) and Yoon *et al., Proc. Natl. Acad. Sci. USA* 93: 2071-2076 (1996).

[0065] One method to increase activity of desired gene products is to use "activation mutagenesis" (*see, e.g.*, Hiyashi *et al. Science* 258:1350-1353 (1992)). In this method an endogenous gene can be modified to be expressed constitutively, ectopically, or excessively by insertion of T-DNA sequences that contain strong/constitutive promoters upstream of the endogenous gene. Activation mutagenesis of the endogenous gene will give the same effect as overexpression of the transgenic nucleic acid in transgenic plants.

Alternatively, an endogenous gene encoding an enhancer of gene product activity or expression of the gene can be modified to be expressed by insertion of T-DNA sequences in a similar manner and FE activity can be increased.

[0066] Another strategy to increase gene expression can involve the use of dominant hyperactive mutants of the gene by expressing modified transgenes. For example, expression of a modified FE with a defective domain that is important for interaction with a negative regulator of FE activity can be used to generate dominant hyperactive FE proteins. Alternatively, expression of truncated FE which have only a domain that interacts with a negative regulator can titrate the negative regulator and thereby increase endogenous FE activity. Use of dominant mutants to hyperactivate target genes is described, *e.g.*, in Mizukami *et al., Plant Cell* 8:831-845 (1996).

SUPPRESSION OF FE EXPRESSION

[0067] A number of methods can be used to inhibit gene expression in plants. For instance, antisense technology can be conveniently used. To accomplish this, a nucleic acid segment from the desired gene is cloned and operably linked to a promoter such that the antisense strand of RNA will be transcribed. The expression cassette is then transformed into plants and the antisense strand of RNA is produced. In plant cells, it has been suggested that antisense RNA inhibits gene expression by preventing the accumulation of mRNA which encodes the enzyme of interest, see, e.g., Sheehy *et al.*, *Proc. Nat. Acad. Sci. USA*, 85:8805-8809 (1988), and Hiatt *et al.*, U.S. Patent No. 4,801,340.

[0068] The nucleic acid segment to be introduced generally will be substantially identical to at least a portion of the endogenous embryo-specific gene or genes to be repressed. The sequence, however, need not be perfectly identical to inhibit expression. The vectors of the present invention can be designed such that the inhibitory effect applies to other proteins within a family of genes exhibiting homology or substantial homology to the target gene.

[0069] For antisense suppression, the introduced sequence also need not be full length relative to either the primary transcription product or fully processed mRNA. Generally, higher homology can be used to compensate for the use of a shorter sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and homology of non-coding segments may be equally effective. Normally, a sequence of

between about 30 or 40 nucleotides and about full length nucleotides should be used, though a sequence of at least about 100 nucleotides is preferred, a sequence of at least about 200 nucleotides is more preferred, and a sequence of at least about 500 nucleotides is especially preferred.

[0070] Catalytic RNA molecules or ribozymes can also be used to inhibit expression of embryo-specific genes. It is possible to design ribozymes that specifically pair with virtually any target RNA and cleave the phosphodiester backbone at a specific location, thereby functionally inactivating the target RNA. In carrying out this cleavage, the ribozyme is not itself altered, and is thus capable of recycling and cleaving other molecules, making it a true enzyme. The inclusion of ribozyme sequences within antisense RNAs confers RNA-cleaving activity upon them, thereby increasing the activity of the constructs.

[0071] Another method of suppression is sense suppression. Introduction of expression cassettes in which a nucleic acid is configured in the sense orientation with respect to the promoter has been shown to be an effective means by which to block the transcription of target genes. For an example of the use of this method to modulate expression of endogenous genes see, Napoli et al., *The Plant Cell* 2:279-289 (1990), and U.S. Patents Nos. 5,034,323, 5,231,020, and 5,283,184.

[0072] Generally, where inhibition of expression is desired, some transcription of the introduced sequence occurs. The effect may occur where the introduced sequence contains no coding sequence per se, but only intron or untranslated sequences homologous to sequences present in the primary transcript of the endogenous sequence. The introduced sequence generally will be substantially identical to the endogenous sequence intended to be repressed. This minimal identity will typically be greater than about 65%, but a higher identity might exert a more effective repression of expression of the endogenous sequences. Substantially greater identity of more than about 80% is preferred, though about 95% to absolute identity would be most preferred. As with antisense regulation, the effect should apply to any other proteins within a similar family of genes exhibiting homology or substantial homology.

[0073] For sense suppression, the introduced sequence in the expression cassette, needing less than absolute identity, also need not be full length, relative to either the primary transcription product or fully processed mRNA. This may be preferred to avoid

concurrent production of some plants which are overexpressers. A higher identity in a shorter than full length sequence compensates for a longer, less identical sequence. Furthermore, the introduced sequence need not have the same intron or exon pattern, and identity of non-coding segments will be equally effective. Normally, a sequence of the size ranges noted above for antisense regulation is used.

PREPARATION OF RECOMBINANT VECTORS

[0074] To use isolated sequences in the above techniques, recombinant DNA vectors suitable for transformation of plant cells are prepared. Techniques for transforming a wide variety of higher plant species are well known and described in the technical and scientific literature. *See*, for example, Weising *et al.*, *Ann. Rev. Genet.* 22:421-477 (1988). A DNA sequence coding for the desired polypeptide, for example a cDNA sequence encoding a full length protein, will preferably be combined with transcriptional and translational initiation regulatory sequences which will direct the transcription of the sequence from the gene in the intended tissues of the transformed plant.

[0075] For example, for overexpression, a plant promoter fragment may be employed which will direct expression of the gene in all tissues of a regenerated plant. Such promoters are referred to herein as "constitutive" promoters and are active under most environmental conditions and states of development or cell differentiation. Examples of constitutive promoters include the cauliflower mosaic virus (CaMV) 35S and 19S transcription initiation regions; the full-length FMV transcript promoter (Gowda *et al.*, *J Cell Biochem* 13D:301; the 1'- or 2'- promoter derived from T-DNA of *Agrobacterium tumefaciens*, and other transcription initiation regions from various plant genes known to those of skill. Such promoters and others are described, *e.g.* in U.S. Patent No. 5,880,330. Such genes include for example, *ACT11* from *Arabidopsis* (Huang *et al.*, *Plant Mol. Biol.* 33:125-139 (1996)), *Cat3* from *Arabidopsis* (GenBank No. U43147, Zhong *et al.*, *Mol. Gen. Genet.* 251:196-203 (1996)), the gene encoding stearoyl-acyl carrier protein desaturase from *Brassica napus* (Genbank No. X74782, Solocombe *et al.* *Plant Physiol.* 104:1167-1176 (1994)), *GPc1* from maize (GenBank No. X15596, Martinez *et al. J. Mol. Biol* 208:551-565 (1989)), and *Gpc2* from maize (GenBank No. U45855, Manjunath *et al.*, *Plant Mol. Biol.* 33:97-112 (1997)).

[0076] Alternatively, the plant promoter may direct expression of a nucleic acid in a specific tissue, organ or cell type (*i.e.*, tissue-specific promoters) or may be otherwise

under more precise environmental or developmental control (*i.e.*, inducible promoters). Examples of environmental conditions that may effect transcription by inducible promoters include anaerobic conditions, elevated temperature, the presence of light, or sprayed with chemicals/hormones. Numerous inducible promoters are known in the art, any of which can be used in the present invention. Such promoters include the yeast metallothionine promoter, which is activated by copper ions (*see, e.g.*, Mett *et al.* (1993) PNAS 90:4567), the dexamethasone-responsive promoter, In2-1 and In2-2, which are activated by substituted benzenesulfonamides, and GRE regulatory sequences, which are glucocorticoid-responsive (Schena *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 88: 0421 (1991)).

[0077] Tissue-specific promoters can be inducible. Similarly, tissue-specific promoters may only promote transcription within a certain time frame of developmental stage within that tissue. Other tissue specific promoters may be active throughout the life cycle of a particular tissue. One of skill will recognize that a tissue-specific promoter may drive expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein a tissue-specific promoter is one that drives expression preferentially in the target tissue or cell type, but may also lead to some expression in other tissues as well.

[0078] In preferred embodiments, promoters that drive fiber-specific expression of polynucleotides can be used. Such expression can be achieved under the control of the fiber-specific promoters described, for example, in U.S. Patent 5,495,070. Typically, the nucleic acids of the invention are operably linked to a promoter active primarily during the stages of cotton fiber cell elongation, *e.g.*, as described by Rinehart (1996) Plant Physiol. 112:1131-1141. See also, John (1997) Proc. Natl. Acad. Sci. USA 89:5769-5773; John, et al., U.S. Patent Nos. 5,608,148 and 5,602,321, describing cotton fiber-specific promoters and methods for the construction of transgenic cotton plants.

[0079] Additional promoters which are linked to genes found to be expressed preferentially in cotton fiber cells can also be identified and isolated for incorporation into the expression cassettes and vectors of the invention . They care also used to express ABP nucleic acids in a cotton fiber specific (or fiber-preferential) manner. As the coding sequences for these tissue specific genes have been characterized, identification and isolation of these cotton fiber specific promoters can be accomplished using standard genetic engineering techniques. For example, Shimizu (1997) Plant Cell Physiol. 38:375-378, found that both endo-1,4-beta-glucanase and expansin mRNA levels were high

during cotton fiber cell elongation, but decreased when cell elongation ceased. Xyloglucan also decreased. The endo-1,3-beta-glucanase mRNA level was very low in the elongating cells, but increased gradually at the onset of secondary wall synthesis, accompanying the massive deposition of cellulose. Also, as discussed above, Song (1997) *supra*, found a cotton fiber-specific acyl-carrier protein in *Gossypium hirsutum* . Ma (1997) Biochim. Biophys. Acta 1344:111-114, found a cotton fiber-specific cDNA encoding a lipid transfer protein . See also John, U.S. Patent No. 5,597,718, describing means to identify cotton fiber-specific genes by differential cDNA library screenings.

[0080] Root-specific promoters may also be used in some embodiments of the present invention. Examples of root-specific promoters include the promoter from the alcohol dehydrogenase gene (DeLisle *et* al. Int. *Rev. Cytol.* 123, 39-60 (1990)).

[0081] Further examples include, *e.g.*, ovule-specific, embryo-specific, endosperm-specific, integument-specific, seed coat-specific, or some combination thereof. A leaf-specific promoter has been identified in maize, Busk (1997) Plant J. 11:1285-1295. The ORF13 promoter from *Agrobacterium* rhizogenes exhibits high activity in roots (Hansen (1997) supra). A maize pollen-specific promoter has been identified, Guerrero (1990) Mol. Gen. Genet. 224:161-168). A tomato promoter active during fruit ripening, senescence and abscission of leaves and, to a lesser extent, of flowers can be used (Blume (1997) Plant J. 12:731-746); or a pistil-specific promoter from the potato SK2 gene, encoding a pistil-specific basic endochitinase (Ficker (1997) Plant Mol. Biol. 35:425-431). The Blec4 gene from pea is active in epidermal tissue of vegetative and floral shoot apices of transgenic alfalfa, making it a useful tool to target the expression of foreign genes to the epidermal layer of actively growing shoots or fibers. Another tissue-specific plant promoter is the ovule-specific BEL1 gene (Reiser (1995) Cell 83:735-742, GenBank No. U39944). See also Klee, U.S. Patent No. 5,589,583, describing a plant promoter region is capable of conferring high levels of transcription in meristematic tissue and/or rapidly dividing cells.

[0082] One of skill will recognize that a tissue-specific promoter may drive expression of operably linked sequences in tissues other than the target tissue. Thus, as used herein a tissue-specific promoter is one that drives expression preferentially in the target tissue, but may also lead to some expression in other tissues as well.

[0083] If proper polypeptide expression is desired, a polyadenylation region at the 3'-end of the coding region should be included. The polyadenylation region can be derived from the natural gene, from a variety of other plant genes, or from T-DNA.

[0084] The vector comprising the sequences (*e.g.*, promoters or coding regions) from genes of the invention will typically comprise a marker gene that confers a selectable phenotype on plant cells. For example, the marker may encode biocide resistance, particularly antibiotic resistance, such as resistance to kanamycin, G418, bleomycin, hygromycin, or herbicide resistance, such as resistance to chlorosulfuron or Basta.

PRODUCTION OF TRANSGENIC PLANTS

[0085] DNA constructs of the invention may be introduced into the genome of the desired plant host by a variety of conventional techniques. For example, the DNA construct may be introduced directly into the genomic DNA of the plant cell using techniques such as electroporation and microinjection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using ballistic methods, such as DNA particle bombardment.

[0086] Microinjection techniques are known in the art and well described in the scientific and patent literature. The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski *et al. Embo. J.* 3:2717-2722 (1984). Electroporation techniques are described in Fromm *et al. Proc. Natl. Acad. Sci. USA* 82:5824 (1985). Ballistic transformation techniques are described in Klein *et al. Nature* 327:70-73 (1987).

[0087] Alternatively, the DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The virulence functions of the *Agrobacterium tumefaciens* host will direct the insertion of the construct and adjacent marker into the plant cell DNA when the cell is infected by the bacteria. *Agrobacterium tumefaciens*-mediated transformation techniques, including disarming and use of binary vectors, are well described in the scientific literature. *See*, for example Horsch *et al.*, *Science* 233:496-498 (1984), and Fraley *et al. Proc. Natl. Acad. Sci. USA* 80:4803 (1983) and *Gene Transfer to Plants*, Potrykus, ed. (Springer-Verlag, Berlin 1995).

[0088] Transformed plant cells which are derived by any of the above transformation techniques can be cultured to regenerate a whole plant which possesses the transformed

genotype and thus the desired phenotype such as increased fiber length, strength or fineness. Such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, typically relying on a biocide and/or herbicide marker that has been introduced together with the desired nucleotide sequences. Plant regeneration from cultured protoplasts is described in Evans *et al.*, *Protoplasts Isolation and Culture*, *Handbook of Plant Cell Culture*, pp. 124-176, MacMillilan Publishing Company, New York, 1983; and Binding, *Regeneration of Plants*, *Plant Protoplasts*, pp. 21-73, CRC Press, Boca Raton, 1985. Regeneration can also be obtained from plant callus, explants, organs, or parts thereof. Such regeneration techniques are described generally in Klee *et al. Ann. Rev. of Plant Phys.* 38:467-486 (1987).

[0089] The nucleic acids of the invention can be used to confer desired traits on essentially any fiber producing plants. These plants include cotton plants (*Gossypium arboreum*, *Gossypium herbaceum*, *Gossypium barbadense* and *Gossypium hirsutum*), silk cotton tree (*Kapok*, *Ceiba pentandra*), desert willow, creosote bush, winterfal, balsa, ramie, kenaf, hemp (*Cannabis sativa*), roselle, jute, sisal abaca and flax.

[0090] One of skill will recognize that after the expression cassette is stably incorporated in transgenic plants and confirmed to be operable, it can be introduced into other plants by sexual crossing. Any of a number of standard breeding techniques can be used, depending upon the species to be crossed.

[0091] Using known procedures one of skill can screen for plants of the invention by detecting the increase or decrease of an mRNA or protein of interest in transgenic plants. Means for detecting and quantifying mRNAs or proteins are well known in the art.

ASSESSING FIBER QUALITY

[0092] Fibers produced from the transgenic plants transformed with FE nucleic acids are compared to control fibers (*e.g.*, fibers from native plants or plants transformed with marker nucleic acids) to determine the extent of modulation of fiber properties. Modulation of fiber properties, such as fiber length, strength, or fineness, is achieved when the percent difference in these fiber properties of transgenic plants and control plants is at least about 10%, preferably at least about 20%, most preferably at least about 30%.

[0093] Several parameters can be measured to compare the properties or quality of fibers produced from transgenic plants transformed with FE nucleic acids and the quality

of fibers produced from native plants. These include: 1) fiber length; 2) fiber strength; and 3) fineness of fibers.

[0094] A number of methods are known in the art to measure these parameters. *See, e.g.*, U.S. Patent 5,495,070, incorporated herein by reference. For example, instruments such as a fibrograph and HVI (high volume instrumentation) systems can be used to measure the length of fibers. The HVI systems can also be used to measure fiber strength. Fiber strength generally refers to the force required to break a bundle of fibers or a single fiber. In HVI testing, the breaking force is expressed in terms of "grams force per tex unit." This is the force required to break a bundle of fibers that is one tex unit in size. In addition, fineness of fibers can be measured, *e.g.*, from a porous air flow test. In a porous air flow test, a weighed sample of fibers is compressed to a given volume and controlled air flow is passed through the sample. The resistance to the air flow is read as micronaire units. More specifically, the micronaire value is a measurement of cotton fiber quality that is a reflection of both fineness and maturity; low values indicate fine and/or immature fiber; high values indicate coarse and/or mature fibers. These values are determined according to standard techniques by measuring the resistance offered by a plug of cotton to airflow (*supra*) that is influenced by a combination of fineness and maturity. Fineness is the outside diameter of the fiber that is measured in mTex (Millitex or mg/km). Maturity is the degree of wall thickening of the fiber. Short fiber count (w) % refers to the percentage of short fiber weight; short fiber count (n) % refers to the percentage of short fiber yield. Immature fiber count refers to the number of immature fibers, *i.e.*, fibers in which the thickening of the fiber wall is appreciably less than normal. An increase in fiber yield (fiber weight/seed) can be measured by using the Advanced Fiber Information System (AFIS). Using these and other methods known in the art, one of skill can readily determine the extent of modulation of fiber characteristics, quality and/or yield in transgenic plants.

EXAMPLES

[0095] The following examples are offered to illustrate, but not to limit the claimed invention.

Example 1

[0096] A full-length cDNA encoding a fiber-PEPC was isolated from a cotton fiber cDNA library. The characteristic features of the complete cDNA clone, designated PEPC1, are described below. PEPC1 gene expression in cotton was evaluated by RT-PCR using total RNA and clone specific primers (Table 1). *PEPC1* was expressed in fibers, embryos, and roots, but not in the light-exposed tissues of leaves, bracts, petals, or stigma. The expression of PEPC1 was higher in fibers collected between 10 to 15 days post anthesis (DPA) when compared to the ovule, embryo and root tissues.

[0097] Developmental changes of PEPC1 expression in fiber (5 to 25 DPA) paralleled the rate of fiber elongation and peaked around the day of maximum cell expansion. Cotton PEPC1 showed a higher level of amino acid identity to the C3-type enzyme from plant species (86-89% identity and 5-7% similarity) compared to 76-86% identity to the C4-type enzyme. These observations suggest that the PEPC1 gene is tissue-specific and can be categorized as a C3-dark PEPC isoform.

[0098] The partial cotton PEPC cDNA clone of 577 bp encodes the C-terminal portion of the enzyme. Sequence comparison of PEPC1 and PEPC2 cDNAs showed 80% nucleotide identity, and the deduced amino acid sequences revealed 87% amino acid identity and 6.3% similarity. Overlapping sequence information confirmed that PEPC2 represents a second highly homologous PEPC gene expressed in developing cotton fibers (*Gossypium hirsufum L.*).

Cloning Techniques:

[0099] A λunizap II cDNA library, constructed from developing cotton fibers (10 days post-anthesis), was screened with a 577 by PCR product as a homologous hybridization probe. The PEPC2 cDNA probe was prepared by PCR-amplification with the primer-set, COT189 (5=-CCATGGATCTTTGCCTGGAC-3=) and COT185 (5=-GCATTCCAGCA GCAATACC-3=), which were designed to conserved regions of PEPC in other plant organisms. The partial cDNA was cloned into pCRII. Both strands of the partial cDNA and the full-length cDNA in pBluescript SK (-) were sequenced by the dideoxy-chain termination method using vector-specific and synthetic oligonucleotide primers.

[0100] Comparison of the nucleotide and the deduced amino acid sequences to alfalfa (AC M83086), Flaveria australasica (AC Z25853), Flaveria pringlef (AC X64144),

Flaveria trinervia (AC X64143), maize (AC X61489), potato (AC X67053), rape seed (AC D13987), sorghum (AC X65137), sorghum (AC X59925), soybean (AC D10717), and tobacco (AC X59016) homologs.

[0101] The PEPC1 cDNA is 3405 bp in length, containing 288 bp 5=-UTR, 2898 bp coding region, and 219 by 3=-UTR. Translation start and stop codons are positioned at nucleotides 289 and 3184, respectively. 44.7% (G+C) content in protein coding region.

[0102] The 2898-by open reading frame of PEPC1 encodes a polypeptide of 965 amino acids with a predicted Mr of 110,183, an isoelectric point of 5.95, and an aliphatic index of 88.54. The total numbers of negatively charged (ASP + GLU) and positively charged (ARG + LYS) residues are 143 and 127, respectively. There is a conserved phosphorylation site at residue 11 (Ser). His-172 and Lys-600, which are essential for enzyme activity, are located in conserved motifs (Chollet et al., 1996). The deduced amino acid sequence exhibits a 79-89% identity and a 5.3-6.9% similarity to the PEPC of other plants.

Protein Homology:

[0103] Alignment of PEPC amino acids (data not shown) reveals some highly conserved regions between these enzymes, which are distributed along the protein. A glycine-rich motif, FHGRGGXXGRGG, found in all PEPC enzymes, is proposed to be involved in the binding of PEP (Iterada et al., 1992). Another highly conserved sequence, GYSDSXKDXG, contains the lysine residue implicated in the activity of PEPC (Jiao et al., 1990). The VXTAHPT motif (amino acid 168-174) contains a histidine residue, which has been suggested to be essential for activity (Chollet et al., 1996; Andreo et al., 1987). Furthermore, a sequence closely related to the phosphorylation site, E/DK/R-X-X-SIDAQLR (Jiao et al., 1990), which is conserved in C3 and C4 plants, is also found in the N-terminal region of the cotton PEPC. It is suggested that non photosynthetic PEPCs from C3 and C4 plants undergo regulatory phosphorylation similar to their C4 and CAM photosynthetic counterparts (Lepiniec et al., 1994).

Expression Characteristics:

[0104] The expression of PEPC in different tissues (fiber, embryo, root, leaf, bract, petal and stigma) was tested by quantitative RT-PCR using 500 ng total RNA and a set of clone-specific primers, COT 285 (5=-CTTTCTGCGGAGTACGAAG-3=) and COT 288

(5=-CCTGCTCTCATCTCATCTTG-3=). PEPC1 was expressed in fibers, embryos, and roots but not in the light-exposed tissues of leaves, bracts, petals, or stigma. The developmental expression of PEPC1 in fiber peaked around the day of maximum fiber elongation.

Example 2

[0105] A fiber cDNA (GhEXP2) (SEQ ID NO: 3) encoding an expansin protein (SEQ ID NO: 4) was isolated from a cotton fiber cDNA library and sequenced. Expansin plays an important role in cell elongation and was targeted for ectopic expression as a potential means for improving traits. The results showed significant gains in cotton yield and all fiber quality properties and provided evidence that cotton plants can be successfully manipulated in order to alter specific traits in a desired direction.

Cotton Fiber Development:

[0106] Scanning electron micrographs have shown that fibers begin to grow and elongate coincident with opening of the flower on the day of anthesis. Fiber growth in the first few days post-anthesis (dpa) has been shown to occur via a diffuse-growing mechanism (Tiwari and Wilkins (1995) *Can. J. Bot.* 73:746-757). By 5 dpa, fibers begin to undergo a period of sustained and rapid cell elongation for 15-20 days, reaching maximum rates of elongation >2mm/day. A proprietary Cotton Fiber dbEST (*i.e.*, a data base with cotton fiber expressed sequence tags), containing about 43,000 sequences, was developed. By using this database it was determined that the genetic complexity in rapidly elongating cotton fibers is high, *i.e.*, requiring about 13,000 unique gene sequences. The fiber genes expressed during fiber elongation represent more than 30% of the cotton genome. It was also determined that expansin is among one of the most abundant gene transcripts expressed in elongating cotton fibers.

Role of Expansin:

[0107] The rigid primary cell wall of plant cells consists of cellulose microfibrils embedded in a matrix of non-cellulose polysaccharides. Cell expansion is driven by turgor pressure and the coordinate regulation of cell wall relaxation. Expansins are a family of cell wall proteins that act to relax the bonds between the cellulose microfibrils and cell wall polymers to allow sufficient slippage and deposition of newly synthesized cell wall materials. The differentially regulated members of the expansin gene family are

associated with auxin-mediated cell expansion in various cell- and tissue-types in dicots. Hence, expansin provided an attractive target for altering cell size and shape.

Vector Construction:

[0108] A 35S::GhEXP2 vector construct was made by employing standard molecular biology recombinant DNA techniques (Sambrook et al. (2000) Molecular Cloning: A Laboratory Manual (Third Edition); Cold Spring Harbor Laboratory Press). A binary vector which includes an enhanced 35S promoter was used for the vector construct and made according to standard laboratory techniques. The cDNA sequence for expansin, GhEXP2, was isolated from a cotton fiber cDNA library (a λunizap II cDNA library was constructed from developing *G. hirsutum* cotton fiber at 10 days post-anthesis). GhEXP2 was cloned into the binary vector and then purified according to standard laboratory techniques (*supra*).

[0109] The GhEXP2 cDNA is 1153 bp in length and encodes an expansin polypeptide of 258 amino acids. A sequence alignment of the expansin polypeptide from GhEXP2 with alpha-expansin precursors (see NCBI, GI:21314545, GI:2134543) and GhEXP1 (see Orford *et al., supra,* GI:2811278) exhibits about 84-86% amino acid identity.

Ectopic Overexpression of Expansin in Transgenic Cotton:

[0110] The 35S::GhEXP2 vector construct was introduced into cotton by standard *Agrobacterium*-mediated co-cultivation of seedling hypocotyls (Agro strain LBA4404). Transgenic cotton plants were selected on kanamycin. Regeneration of T0 kanamycin-resistant plants was accomplished via somatic embryogenesis in about 10-12 month. Selection of fertile, kanamycin-resistant plants was made following confirmation of the presence of the intact transgene at the molecular level.

Fiber Analysis:

[0111] Cotton seed (*i.e.*, seed and fiber) was harvested from mature bolls grown in the greenhouse. Fiber and seed were weighed. Then fiber was removed from the seed by hand. Fiber and seed were again weighed separately which provided data for some of the yield components. The fiber was shipped to a fiber analysis lab (Cotton Inc.) where fiber properties were measured using standard procedures. The measurements were performed by using the Advanced Fiber Information System (AFIS) as employed in the art.

Results:

[0112] The results showed significant gains in yield and fiber quality of T1 transgenic cotton plants that overexpress expansin. Wild type cotton plants (control) were compared to T0 kanamycin-resistant plants and T1 transgenic plants. The T1 transgenic plants showed a 4 fold increase in yield (fiber weight/seed); an increase in fiber length by about 0.2 inch (gains measured in 0.01 inch); a 50% decrease in immature fibers *(i.e.*, increased number of mature fibers); and a 50% decrease in short fibers (*i.e.*, increased fiber uniformity).

[0113] There was a dramatic increase in the number of fiber initials in overexpressing expansin transgenic plants relative to wild type control (untransformed) plants as determined through scanning electron micropgraphs. The length of fiber initials in transgenic plants at 0 dpa (anthesis) was also increased by at least a factor of 2. Thus, increased yield in transgenic plants is due in part to an increase in fiber number/seed.

[0114] As shown above, ectopic overexpression of a single gene encoding the cell wall protein expansin successfully and significantly improved complex agronomic traits such as yield and fiber quality.

[0115] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference in their entirety for all purposes.

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule comprising a FE polynucleotide sequence, which is at least about 90% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO: 5, SEQ ID NO: 7, or SEQ ID NO: 9.

2. The isolated nucleic acid molecule of claim 1, wherein the isolated nucleic acid molecule comprises at least about 100 nucleotides.

3. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid encodes an expansin polypeptide having a sequence of SEQ ID NO:4.

4. The isolated nucleic acid molecule of claim 1, wherein the nucleic acid is an expansin polynucleotide as shown in SEQ ID NO:3.

5. The isolated nucleic acid molecule of claim 1, further comprising a plant promoter operably linked to the FE polynucleotide.

6. The isolated nucleic acid molecule of claim 5, wherein the plant promoter is a tissue-specific promoter.

7. The isolated nucleic acid molecule of claim 6, wherein the promoter is a fiber-specific promoter.

8. The isolated nucleic acid molecule of claim 5, wherein the promoter is an inducible promoter.

9. The isolated nucleic acid molecule of claim 5, wherein the FE polynucleotide is linked to the promoter in an antisense orientation.

10. A transgenic plant comprising an expression cassette containing a plant promoter operably linked to a heterologous FE polynucleotide sequence, which is at least about 90% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, and SEQ ID NO:9.

11. The transgenic plant of claim 10, wherein the FE polynucleotide encodes an FE polypeptide.

12. The transgenic plant of claim 11, wherein the FE polypeptide has an amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, or SEQ ID NO:10.

13. The transgenic plant of claim 10, wherein the heterologous FE polynucleotide is linked to the promoter in an antisense orientation.

14. The transgenic plant of claim 10, wherein the plant promoter is a tissue specific promoter.

15. The transgenic plant of claim 14, wherein the promoter is a fiber-specific promoter.

16. The transgenic plant of claim 15, wherein the promoter is a cotton fiber specific promoter.

17. The transgenic plant of claim 10, wherein the FE polynucleotide has a nucleic acid sequence of SEQ ID NO:3.

18. The transgenic plant of claim 10, wherein the plant is a member of the genus *Gossypium*.

19. A method of modulating fiber quality and yield in a plant, the method comprising introducing into the plant an expression cassette containing a plant promoter operably linked to a heterologous FE polynucleotide sequence, which is at least about 90% identical to SEQ ID NO:1, SEQ ID NO:3, SEQ ID NO:5, SEQ ID NO:7, and SEQ ID NO:9.

20. The method of claim 19, wherein the heterologous FE polynucleotide encodes a FE polypeptide.

21. The method of claim 20, wherein the FE polypeptide has an amino acid sequence of SEQ ID NO:2, SEQ ID NO:4, SEQ ID NO:6, SEQ ID NO:8, or SEQ ID NO:10.

22. The method of claim 19, wherein the heterologous FE polynucleotide is linked to the promoter in an antisense orientation.

23. The method of claim 19, wherein the heterologous FE polynucleotide has a nucleic acid sequence of SEQ ID NO:3.

24. The method of claim 19, wherein the plant promoter is a tissue specific promoter.

25. The method of claim 24, wherein the promoter is a fiber specific promoter.

26. The method of claim 25, wherein the promoter is a cotton fiber specific promoter.

27. The method of claim 19, wherein the plant is a member of the genus *Gossypium*.

28. The method of claim 19, wherein the expression cassette is introduced into the plant through a sexual cross.

29. The method of claim 19, wherein the heterologous FE polynucleotide encodes an expansin polypeptide.

SEQUENCE LISTING

SEQ ID NO:1

PEPCase: Cotton +10 dpa Fiber cDNA

GenBank Accession Number AF008939 (PEPC1)

TTTCACTTTCTGGTGTTCAGTCATAAGTTACCCATTATTTTATCTGTTTTTCTAGCGATTTCTTGGT
GTGGGGCATCACGGAATTCACTTCCAAAGAGATAGGATTTGATTCATCTTTTGTTTTCCGGGTTTG
GGTTTGGGTTCGGCGGATTCTTCGCCACAAATTATTTTTATCCTTTTACACCGCAAAATAATAAAT
TTCATCCGTAATTGTTCCTTTCTTTTTCCTGCCTTTCATTGGTTCATAATTTAAAGTTTTGGCTTTGG
TAGTCAAGTAAAAAGACAGAGATGGCAGGTAGAAAAGTAGAGAAGATGGCTTCAATAGATGCA
CAACTAAGGCTATTAGCACCTGGGAAAGTCTCTGAGGATGATAAGTTGGTTGAATATGATGCTGT
GCTACTGGATCGTTTTCTTGATATTCTTCAGGATTTGCATGGAGAGGATATCAGAGAAACGGTTC
AAGAATGCTACGAGCTTTCTGCGGAGTACGAAGGAAAGCATGATCCTAAAATATTGGAGGAGCT
TGGCAAGGTACTTACAAGCTTGGATCCTGGGGACTCAATTGTTGTTACCAAATCATTTTCCCACAT
GCTTAACTTGGGAAACTTGGCCGAGGAGGTTCAAATTGCATATAGACGAAGGATTAAGTTGAAG
AAAGGAGATTTTGCTGATGAGAGTTCAGCAACAACTGAATCAGATATTGAAGAGACGTTCAAGC
GACTTGTGGGGCAGTTGAACAAGTCCCCTGAAGAAGTTTTTGATGCTTTGAAGAATCAGACTGTA
GATTTAGTCTTAACTGCACATCCTACACAGTCTGTTCGAAGATCTTTACTTCAAAAACACGGAAG
GATACGTAATTGTTTGACACAATTGTATGCTAAAGACATTACACCTGATGATAAGCAGGAACTTG
ATGAGGCACTACAAAGGGAGATTCAAGCTGCTTTTCGCACAGATGAGATCCGAAGGAATCCTCC
CACCCCACAAGATGAGATGAGAGCAGGAATGAGCTACTTCCACGAGACAATCTGGAAAGGTGTA
CCAAAATTCTTGCGTCGTGTTGACACAGCTTTGAAGAACATAGGGATAAATGAACGTGTTCCATA
CAATGCCCCTCTTATTCAATTCTCTTCATGGATGGGAGGGGATCGTGATGGAAACCCCAGGGTAA
CTCCTGAAGTTACAAGAGATGTCTGCTTATTAGCTAGAATGATGGCTGCTAACTTGTACTTCTCCC
AAATAGAGGATCTTATGTTTGAGTTATCAATGTGGCGTTGCAGCGATGAACTTCGTATTCGTGCA
GATGAACTCCATAGGTCCTCCAAAAAGGATGCAAAGCATTACATAGAATTTTGGAAACAGATTCC
TCCAAATGAGCCATATCGCATTATTCTTGGTGATGTGAGGGACAAGCTGTATAATACACGTGAAC
GTGCTCGTAGCCTGTTGGCCAATGGATTTTCTGACATTCCTGAAGAAGCAGCATTTACCAATGTG
GAGCAGTTTCTGGAGCCTCTTGAACTCTGCTATAGATCACTCTGTGCTTGTGGTGATCGGCCAATA
GCTGATGGAAGCCTTCTTGATTTCTTACGGCAAGTTTCTACCTTTGGGCTTTCACTTGTGAGGCTT
GATATCCGACAGGAATCTGATAGACATACTGATGTCCTTGATGCTATCACAAAGCACCTGGATAT
TGGATCTTATCGAGAATGGCCTGAGGAACGCCGACAGGAATGGCTCTTATCTGAACTCAGAGGC
AAGCGCCCTCTATTCGGCCCTGATCTTCCCAAAACAGAAGAAGTCGCTGATGTATTGGACACATT
TCATGTCATTTCTGAACTGCCTTCAGACAGCTTTGGTGCCTATATAATCTCAATGGCTACAGCCCC
ATCTGATGTGCTTGCTGTTGAGCTTTTACAACGTGAATGCCATGTAAAGCAACCATTACGGGTAG

TTCCATTGTTTGAAAAACTCGCTGATCTTGAAGCTGCTCCTGCTGCTGTGGCTCGTCTCTTCTCTAT
AGATTGGTACAGAGACCGGATCAATGGGAAACAGGAAGTGATGATAGGTTATTCAGATTCAGGA
AAGGATGCTGGCCGTCTTTCTGCAGCATGGCAGCTATACAAGGCTCAGGAGGAACTTGTAAAGGT
GGCAAAGCAGTATGGTGTTAAGCTTACAATGTTCCATGGCCGAGGAGGGACAGTTGGAAGAGGA
GGGGGACCCACGCATCTTGCTATATTGTCTCAACCACCCGATACAATTCATGGATCACTTCGTGT
AACAGTTCAAGGTGAAGTTATTGAACAATCATTTGGAGAGGAGCACTTGTGCTTCAGGACGCTTC
AACGTTTTACTGCTGCTACACTTGAGCATGGAATGCATCCCCCTGTCTCACCAAATCCAGAATGG
CGTGCACTCATGGATGAAATGGCAGTAGTTGCAACAAAGGAATACCGTTCTGTAGTCTTCCAGGA
ACCTCGCTTTGTTGAATACTTCCGCCTAGCAACACCAGAATTGGAGTATGGTCGGATGAATATTG
GAAGCCGTCCATCAAAAAGGAAACCAAGTGGAGGCATTGAATCACTCCGTGCAATCCCATGGAT
CTTTGCATGGACTCAAACAAGATTTCATTTACCTGTGTGGCTTGGCTTTGGGGCTGCATTTAAGCA
CGTCATTCAGAAGGATATAAAGAATCTCCACATGCTCCAGGAGATGCACAACCAATGGCCTTTCT
TTAGGGTCACAATGGACTTAATTGAAATGGTATTTGCCAAGGGAGACCCTGGAATCGCTGCCTTA
TATGACAAGCTGCTAGTGTCAAAGGAACTCTGGCCCTTTGGAGAGAACTTGAGAGCTAACTATGA
AGACACTAAACGACTTGTTCTCCAGGTTGCTGGACACAGAGATCTTCTTGAAGGTGACCCTTACC
TGAAGCAGAGGCTCCGACTTCGTGATGCTTACATCACAACCCTTAATGTGTGCCAAGCGTACACT
TTGAAAAGGATCCGTGACCCTGATTATCATGTCAAGGTCAGGCCTCACTTATCCAGGGAATACAT
GGAATCAAGCAAGGCGGCAGCTGAGCTTGTGAAACTTAACCCTACAAGCGAGTATGCTCCTGGT
CTGGAAGACACCCTTATTTTGACCATGAAGGGTATTGCTGCTGGAATGCAAAACACTGGTTAAAA
CTTGCGGATTGTTGTTTCCTTGTGTTGTTATGCTCCTTAGTTTATTTTTAATGGATGTTGCTTTGCG
TCAAGTAAACCTTTTATATTACTATTATTATCGGTTGGGTTAATTTCATATTTGGCTCATATATTAC
TAAGTGTTGTCAGTCTGGTATCTCTATAATAATACATATTAACATGATTTCCCATTCAAAAAAAAA
AAAAAAAAAAAAAAAAA

SEQ ID NO:2

PEPcase protein sequence

GenBank Accession Number AF008939 (PEPC1)

MAGRKVEKMASIDAQLRLLAPGKVSEDDKLVEYDAVLLDRFLDILQDLHGEDIRETVQECYELSAEY
EGKHDPKILEELGKVLTSLDPGDSIVVTKSFSHMLNLGNLAEEVQIAYRRRIKLKKGDFADESSATTES
DIEETFKRLVGQLNKSPEEVFDALKNQTVDLVLTAHPTQSVRRSLLQKHGRIRNCLTQLYAKDITPDD
KQELDEALQREIQAAFRTDEIRRNPPTPQDEMRAGMSYFHETIWKGVPKFLRRVDTALKNIGINERVP
YNAPLIQFSSWMGGDRDGNPRVTPEVTRDVCLLARMMAANLYFSQIEDLMFELSMWRCSDELRIRA
DELHRSSKKDAKHYIEFWKQIPPNEPYRIILGDVRDKLYNTRERARSLLANGFSDIPEEAAFTNVEQFL
EPLELCYRSLCACGDRPIADGSLLDFLRQVSTFGLSLVRLDIRQESDRHTDVLDAITKHLDIGSYREWP
EERRQEWLLSELRGKRPLFGPDLPKTEEVADVLDTFHVISELPSDSFGAYIISMATAPSDVLAVELLQR
ECHVKQPLRVVPLFEKLADLEAAPAAVARLFSIDWYRDRINGKQEVMIGYSDSGKDAGRLSAAWQL

YKAQEELVKVAKQYGVKLTMFHGRGGTVGRGGGPTHLAILSQPPDTIHGSLRVTVQGEVIEQSFGEE
HLCFRTLQRFTAATLEHGMHPPVSPNPEWRALMDEMAVVATKEYRSVVFQEPRFVEYFRLATPELEY
GRMNIGSRPSKRKPSGGIESLRAIPWIFAWTQTRFHLPVWLGFGAAFKHVIQKDIKNLHMLQEMHNQ
WPFFRVTMDLIEMVFAKGDPGIAALYDKLLVSKELWPFGENLRANYEDTKRLVLQVAGHRDLLEGD
PYLKQRLRLRDAYITTLNVCQAYTLKRIRDPDYHVK
VRPHLSREYMESSKAAAELVKLNPTSEYAPGLEDTLILTMKGIAAGMQNTG

SEQ ID NO:3

Expansin cDNA from 10 dpa fiber ZAPII library from G. hirsutum cv. Acala SJ-2

GGCACGAGGGTCAGCCAATTGTTTGAGCTAGCTAGCTCTTACTCAAATGGCAACCAAAACGATGA
TGTTGCAAATATTTTCACTTTTCTTCTTTTTGTTCAGTATCTGCAACTCCATTTTCCTTGGTGCTAAT
GGAGATGACAATGGTGGTTGGCAAACTGCCCATGCCACCTTCTACGGTGGTGCTGATGCTACCGG
CACAATGGGGGGGAGCTTGTGGTTATGGAAACCTGTACAGTCAAGGGTATGGAACGAGCACAGCA
GCTTTGAGCACTGCACTTTTCAACAATGGCTTGAGCTGCGGTGCCTTCTACGAGCTCCGGTGCAA
CAATGATCCTCAATGGTGCATTAGTCGAACCATAACCGTGACAGCCACCAACTTTTGTCCCCCTA
ACTATGCTTTATCTAGTGACAATGGCGGGTGGTGCAATCCCCCACGAGAACACTTTGATTTGGCC
GAACCGGCATTCTTGCAGATCGCGGAATATCGAGCTGGGATCGTCCCTGTTATGTTCAGAAGGGT
GTCATGTGTGAAGAAAGGAGGCATCAGGTACACCATGAATGGACATTCGTACTTCAACATGGTGT
TGATAACCAACGTGGGAGGGGCAGGGGATATAACGTCAGTGTCCATCAAGGGTTCCAAAACAGG
ATGGCTACCTATGTCCAGAAATTGGGGCCAAAACTGGCAGAGCAATGCTTACCTTAACGGCCAA
AGCCTCTCTTTCAAAGTGACTGCCAGCGATAGCAGGACTATCACAAACTACAATGTAGTGCCCGC
TGGTTGGCAATTCGGACAAACTTTTGAAGGAGGCCAGTTTTAAGACAATATTATAGTGTCCGTCT
AATATTAAAACTGGAATTGACATATTACTTATATAAGGCACATGAGCGTTTTATGCCGAGGTAGC
AAAGTGGCGCCCGCTGGCTTTATGTGTGAAATAGGCGAGCAAGTGCCATTAGCCTATAATATACA
CATTTCCTATAGTGAACCAAACTATTAAGTTTGAACTCTAGAAGTGATATCCATAATGTCTGAAA
TTTGATTGTTGATGATTGACCATGATATTTATGGTTTTCATTATTGAAATACTTTTTTATTATAATT
TATAAATAAATGAATCATTTCTTTTTACAAAAAAAAAAAAAAAAAA

SEQ ID NO:4

Expansin protein sequence

MATKTMMLQIFSLFFFLFSICNSIFLGANGDDNGGWQTAHATFYGGADATGTMGGACGYGNLYSQG
YGTSTAALSTALFNNGLSCGAFYELRCNNDPQWCISRTITVTATNFCPPNYALSSDNGGWCNPPREHF
DLAEPAFLQIAEYRAGIVPVMFRRVSCVKKGGIRYTMNGHSYFNMVLITNVGGAGDITSVSIKGSKTG
WLPMSRNWGQNWQSNAYLNGQSLSFKVTASDSRTITNYNVVPAGWQFG QTFEGGQF

SEQ ID NO:5

Cotton (Gossypium hirsutum L. cv Acala SJ-2) 10 dpa fiber cDNA encoding the Korrigan homolog from Arabidopsis, a member of the EGase family of endo-1,4-beta-D-glucanases

GGCACGAGCCACTTTGGAAAACCAATGAAGAAAGCACCACTGCCTACTTATACACATTTTATTTA
TCAAACTCAACTTCCTCTTCGTCTTGGTAGATCAGATCTGTCGGTGCTCCTGCATTTTCCGCCCAC
TACTCTTCCAAATCCTCATCATGTACGGCAGAAATCCGTGGGGAGGTCCCCTGGAGATAAACGCC
ACTGATTCTGCCACTGACGACGACAGGAGCAGGAATCTGCAGGACCTGGATAGGGCTGCACTCT
CTCGCCCCTTGGACGAGACTCAGCAAAGCTGGCTGCTTGGCCCCGGGGAGCAAAAGAAGAAGAA
GAAGTACGTTGATCTCGGATGTATCATTGTGAGCCGCAAGATCTTTGTATGGACCGTGGGGACCC
TGCTAGTCTCCGCCCTCCTGGCCGGACTCATCACCCTCATCGTCAAGACTGTCCCACGTCATCACC
ACCGCCACTCTCCGCCCGATAACTACACTCTGGCTCTTCACAAGGCGCTCATGTTCTTTAATGCTC
AGCGTTCTGGAAAGCTGCCCAAGCATAATAATGTGTCGTGGAGAGGGAACTCGGGCCTCCAAGA
TGGCAAATCCGATCCCTCCGTTTTGATGAAAGATCTGGTCGGCGGATATTACGATGCTGGAGATG
CTATCAAGTTTAACTTTCCTGCATCTTTTTCAATGACTATGTTGAGCTGGAGTGTCATCGAATACA
GTGCTAAATACGAGGCTGCCGGCGAGCTCAATCATGTTAAAGAGATCATCAAATGGGGTACTGA
TTATCTTCTGAAGACCTTCAACAATACTGCTGATACCATTGACAGGATTGCAGCGCAGGTAGGGA
TAGGAGATACATCTGGAGGAGTTTCAGCCCCAAATGATCATTATTGCTGGATGCGCCCTGAGGAC
ATTGATTACCCCCGTCCTGTATATGAATGTCATAGTTGCTCCGATCTTGCTGCTGAAATGGCTGCT
GCTTTGGCTTCGGCTTCCATCGTTTTCAAAGACAACAAAGCATACTCTCAAAAGCTTGTCCATGGT
GCCCGAACACTCTTTAAGTTTGCTAGGGATCAAAGAGGCAGATATAGTGCTGGTCGTTCTGACCC
TGCCCTCTTTTATAATTCCTCAAGTTACTGGGATGAGTTTGTTTGGGGTGGAGCCTGGTTATACTA
TGCCACTGGGAATTCATCCTATCTTCAGTTAGCTACTCATCCTAAACTTGCCAAGCATGCTGGTGC
TTTCTGGGGTGGCCCAGATTATGGTGTTCTTAGCTGGGATAATAAGCTTGCTGGTGCTCAGGTGCT
TCTGAGCCGATTGAGATTGTTTTTGAGTCCTGGGTATCCATATGAGGAAATATTGAGTACGTTTCA
TAATCAAACCAGCATAATTATGTGCTCATTCCTTCCGGTTTTCACTAGCTTTAATAGAACAAAAGG
AGGTTTGATTCAGTTAAACCATGGAAGGCCTCAGCCACTGCAATACGTAGTCAATGCAGCCTTCT
TAGCCGCCCTATATAGTGATTATCTTGATGCTGCTGATACACCTGGATGGTATTGTGGTCCCAATT
TCTATTCAACTGATGTCCTGCGTGAATTTGCCAAAACCCAGATTGATTATATCCTTGGCAAAAATC
CTCGAAAAATGAGCTATGTTGTGGGCTTTGGTAACCATTATCCAAAGCATGTTCACCATAGAGGG
GCATCTATCCCTAAGAATAAGATCAAATATAACTGTAAAGGGGGATGGAAATGGAGGGATACGT
CAAAACCAAACCCCAACACACTTGTGGGAGCCATGGTAGCAGGACCTGACAAGCATGATGGGTT
TCGTGATGTTCGCACCAACTACAACTATACGGAGCCAACTCTAGCAGGCAACGCAGGGTTGGTTG
CTGCACTCGTGGCATTGTCTGGTGACAAGGCAACCGTGATTGACAAGAATACTATTTTTTCTGCA
GTTCCACCAATGTTTCCTACACCACCACCACTTCCGGCACCTTGGAAACCATGAAAACGTTTTGAT
CTTTCTTCTGTCCATGTGTGACTTACAGTCTGATGATTTTGGAATTAGTTTTTGGTACGTAAATGA

CCTTGGAAGTGTAAGTAACGCAAAAAGGCAAGACAGGAGATGAGTGATATGAGACAACCTGGAC
AAAACTCGTGGTGTTTGGCTGTCAGATTGGAGAGTCAGAGAAGCATTCAAACATTGATAATTTGT
ATATGAAGTCTATTTGGATTTTGTGGTATACAGACAGACGGAGTGCATGTATGATATTAATGTTG
TATTATTTGAAACACACCAATATCCATTGACAATAAATTTGTGTTTAATTGTTGCAGAAATCTGGC
CTTTGCTATTACCAAAAAAAAAAAAAAAAAACTCGAG

SEQ ID NO:6

Cotton (Gossypium hirsutum L.) 10 dpa Fiber homolog of the Arabidopsis Korrigan,
a member of the endo-1,4-beta-D-glucanase(EGase)family

ARATLENQRKHHCLLIHILFIKLNFLFVLVDQICRCSCIFRPLLFQILIMYGRNPWGGPLEINATDSATD
DDRSRNLQDLDRAALSRPLDETQQSWLLGPGEQKKKKKYVDLGCIIVSRKIFVWTVGTLLVSALLAG
LITLIVKTVPRHHHRHSPPDNYTLALHKALMFFNAQRSGKLPKHNNVSWRGNSGLQDGKSDPSVLMK
DLVGGYYDAGDAIKFNFPASFSMTMLSWSVIEYSAKYEAAGELNHVKEIIKWGTDYLLKTFNNTADT
IDRIAAQVGIGDTSGGVSAPNDHYCWMRPEDIDYPRPVYECHSCSDLAAEMAAALASASIVFKDNKA
YSQKLVHGARTLFKFARDQRGRYSAGRSDPALFYNSSSYWDEFVWGGAWLYYATGNSSYLQLATH
PKLAKHAGAFWGGPDYGVLSWDNKLAGAQVLLSRLRLFLSPGYPYEEILSTFHNQTSIIMCSFLPVFT
SFNRTKGGLIQLNHGRPQPLQYVVNAAFLAALYSDYLDAADTPGWYCGPNFYSTDVLREFAKTQIDY
ILGKNPRKMSYVVGFGNHYPKHVHHRGASIPKNKIKYNCKGGWKWRDTSKPNPNTLVGAMVAGPD
KHDGFRDVRTNYNYTEPTLAGNAGLVAALVALSGDKATVIDKNTIFSAVPPMFPTPPPLPAPWKPKRF
DLSSVHVLTVFWNFLVRKPWKCKRKKARQEMSDMRQPGQNSWCLAVRLESQRSIQTLIICISLFGFC
GIQTDGVHVYCCIINTPISIDNKFVFNCCRNLAFAITKKKKKKLE

SEQ ID NO:7

XET

Full-length Gossypium hirsutum L. (cotton) fiber gene isolated from a 10 dpa cDNA library.

CTCAAACCACCTTGTTCTCCCCCCTTCTCCGTCTTTCTGGCAACATGGGTTCACCACCACTTTGGG
TTCTGCTTCTGGGTGTGTTGTTTATGGCTTCCGGAACAATCGCAGCTCCCCCTAAGAAGCCTGTAG
ATGTACCATTCTCCAGAAACTATATGCCTACTTGGGCTTTTGATCACATTAAGTATTTCAATGGTG
GCTCTGACATTCAGCTCCACCTTGACAAATACACTGGTACTGGTTTCCAGTCCAAAGGATCATAC
TTGTTCGGACACTTCAGTATGCAAATAAAGCTAGTCCCTGGAGATTCTGCTGGGACTGTCACTGC
CTGTTATTTGTCTTCTCAAAACTCAGAGCATGATGAGATAGATTTTGAGTTCTTGGGCAACAGAA
CAGGGCAACCATACATTCTTCAGACCAATGTGTTCACTGGTGGCAAAGGAGACAGAGAACAAAG
GATTTACCTTTGGTTTGACCCAACCAAAGAATACCACTCCTACTCTGTCCTCTGGAACATGTATCA
GATANTGTTCTTTGTGGACGACATACCAATCAGAGTGTTCAAAAACTGCAAAGATTTGGGAGTGA
GGTCAAGGGGTGGTCTCGAAAAGACCGACTGGTCCAAAGCCCCATTCATAGCCGCATACAAGAG
CTTCCACATCGACGGGTGCGAGTCGTCGGTGGAAGCCAAGTTCTGCGCCACACAGGGAAAGCGG

TGGTGGGACCAAAAGGCATTCGAGGACCTCGACGCCTATCAGTGGCGCAGACTGCGTTGGGTCC
GCAACAAGTTCACTATTTACAACTATTGCAGCGATAGGGTGAGGTACCCCACAATGTCGCCCGAG
TGCAAGAGAGACAGAGACGCTTGAGTCATTTCCATCACCACCGAGTCAATTGCTTGTTGGCCTTC
ACTTATTTCCACATCAATTTCATTATATGTAATTTCACTCCTAAAACATTTGTTTTCGTGATAATAT
TATTTAAACATAGATTACCATCTCTTGACGGGGGACAAGACCATNATCTTGTAAGTATAAGTATA
AGAAGCTTGGGATTCGATTTCAAGAAATNNAAGTCTATTTAAAAAAAAAAAAAAAAAAAACTCG
AGGGGGGGCCCGGTACCCAATTCCGCCCTATAGTTGAGTCCTATACNATTCCACTGGGCCGTCTT
TTTACAAC

SEQ ID NO:8

XET

Deduced amino acid sequence of Gossypium hirsutum L. (cotton) fiber cDNA isolated from a 10 dpa fiber cDNA library

MGSPPLWVLLLGVLFMASGTIAAPPKKPVDVPFSRNYMPTWAFDHIKYFNGGSDIQLHLDKYTGTGF
QSKGSYLFGHFSMQIKLVPGDSAGTVTACYLSSQNSEHDEIDFEFLGNRTGQPYILQTNVFTGGKGDR
EQRIYLWFDPTKEYHSYSVLWNMYQIXFFVDDIPIRVFKNCKDLGVRSRGGLEKTDWSKAPFIAAYKS
FHIDGCESSVEAKFCATQGKRWWDQKAFEDLDAYQWRRLRWVRNKFTIYNYCSDRVRYPTMSPEC
KRDRDA*VISITTESIACWPSLISTSISLYVISLLKHLFS**YYLNIDYHLLTGDKTXIL*V*V*EAWDSISR
NXSLFKKKKKKNSRGGPVPNSAL*LSPIXFHWAVFLQ

SEQ ID NO:9

Consensus PME contig sequence

This is a consensus sequence for G. hirsutum L. Pectin methylesterase cotton fiber (~10 dpa) gene assembled from nucleotide sequences 5PME, 3PME, 2PME, and the Gossypium arboreum L. cv. AKA8401 Efiber EST xgi contig CON_001_18878.

AAAGCTTTTCTTGGCTCTCTTTGCGTCAATCCTTCTTGTAACTGCCATAGTCACCATTGCCACCAC
CGTCTCCATTTCCAAAAAGAAATCCAGTAATACTGTAGCAGCTCACTCCATCATCAAATCTTCAT
GTAGCTCCACGTTGTACCCAGAGTTATGCTACTCAACAATCTCTTCAGCACCAGATGCTGAGACC
AAGGTCAAGAACCCCAAGGATGTGATTGAATTGTCGTTGAACTTGACGGTGACTGCTGTTCAGAG
TAACTATTTGTCCATCAAAAAGCTCATTAGTACCCGAAGGAAGAGCCTCACGGAGCGCGAAAAG
GCTGCCCTTAACGATTGTCTTGAACTAGTGGATGAGACTTTGGATGAGCTATTCGTAGCTGAACA
TGATCTCAGTGACTATCCAAGCTTTAACAAGTCAATTTCCCAACATGCTGATGACCTTAAGAGTCT
TCTTAGTGCTGCAATGACCAACCAAGAAACTTGCCTTGATGGGTTTTCTCACGATAAAGCTGATA
AAAAGGTGAGGCAAGCGTTGCTTGACGGGCAGATGCATGTTTTTCATATGTGTAGTAATGCCCTG
GCAATGATCAAGAACTTGACGGACACAGACATGGCAAGCCAAGGTTATCATCCATCATCTGGGA
GGCAACTTGAGGAGCAAGACCAAACAGAATGGCCTAAATGGCTGTCGGAGGGAGATAGGAGAC

TGTTACAGGCTACAACAGTGATTCCTAATGTAACAGTGGCCGCTGATGGTAGTGGAGACTTCCTC
ACGGTGTCTGAGGCGGTGGCGGCTGCACCGGAGAGAAGCACCACGAGGTACATTATTAAGATTA
AAGCTGGAGTTTATAGGGAAACGTGGATGTTTCCAAGTAAGAAAACCAATCTCATGTTTGTGGGA
GATGGGAGGGTCAACACCATCATCACAGCTAGCAGAAATGTTGTCGATGGCAGCACCACTTTCCA
CTCTGCCACTGTTGCTGCGGTAGGGGACGGGTTCTTGGCCAGGGATATAACATTTCAGAACACGG
CTGGACCATCGAAGCACCAAGCAGTGGCACTGCGTGTGGGCTCTGATTTATCAGCATTCTACAGG
TGTGGCATTTTAGCATACCAGGACACTCTCTATGTCCACAGCCTTCGCCAATTCTATTCACAATGC
CTTGTAGCAGGCAGCGTGGACTTCATATTCGGAAATGCAGCAGCAGTGTTGCAAGACTGCGACAT
TCATGCTCGTCGACCCAATCCAAACCAAAGGAACATGGTCACCGCACAAGGNCGTAGTGACCCA
AACGAGAACACTGGGATTGTGATTCANAAATGTNGGATCGGTGCAACCTCGGATTTAGAAGCCG
TTAAATCCGATTTTGAAACTTATTTAGGGAGACCATGGAAGACACATTCGAAGACTGTTATCATG
CAATCTGTTATAAGTGATATTATTCATCCTGCTGGTTGGTTCCCATGGGGAAAAAAAAATTCGCACT
CNACCTTTGACGTATCNGGAATATCANAATACTNGGCCTGGANCTTAACNCNTCAANCAGGGTTA
CATGGAAAAGGGTTATTACNTTATCCCCCACATATCCGGAAGCCCAAAACCTACCTGCCTCCGNA
ATTTTTNTTTGGGGGAACNTAATTGGGTTTANCCCCNCCGGGNCTTNCCTTTCCCTCNTNGAATCT
TTGAAAA

SEQ ID NO:10

PME consensus protein seq

This is the deduced protein sequence for a G. hirsutum L. fiber pectin methyesterase derived from the PME consensus seqeunce.

XXXXXXXXXXKLFLALFASILLVTAIVTIATTVSISKKKSSNTVAAHSIIKSSCSSTLYPE
LCYSTISSAPDAETKVKNPKDVIELSLNLTVTAVQSNYLSIKKLISTRRKSLTEREKAALNDCLELVDET
LDELFVAEHDLSDYPSFNKSISQHADDLKSLLSAAMTNQETCLDGFSHDKADKKVRQALLDGQMHV
FHMCSNALAMIKNLTDTDMASQGYHPSSGRQLEEQDQTEWPKWLSEGDRRLLQATTVIPNVTVAAD
GSGDFLTVSEAVAAAPERSTTRYIIKIKAGVYRETWMFPSKKTNLMFVGDGRVNTIITASRNVVDGST
TFHSATVAAVGDGFLARDITFQNTAGPSKHQAVALRVGSDLSAFYRCGILAYQDTLYVHSLRQFYSQ
CLVAGSVDFIFGNAAAVLQDCDIHARRPNPNQRNMVTAQXRSDPNENTGIVIXKCXIGATSDLEAVKS
DFETYLGRPWKTHSKTVIMQSVISDIIHPAGWFPWGKKIRTXPLTYXEYXNTXPGX*XXXQGYMEKG
YYXIPHISGSPKPTCLXNFXLGEXNWVXPXRXXPFPXXIFEX

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
27 November 2003 (27.11.2003)

PCT

(10) International Publication Number
WO 2003/098186 A3

(51) International Patent Classification[7]: C07H 21/04, C12N 15/82, 15/87, A01H 5/00

(21) International Application Number: PCT/US2003/015269

(22) International Filing Date: 16 May 2003 (16.05.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
10/150,559 17 May 2002 (17.05.2002) US

(71) Applicant *(for all designated States except US)*: **THE REGENTS OF THE UNIVERSTIY OF CALIFORNIA** [US/US]; 1111 Franklin Street, Fifth Floor, Oakland, CA 94607-5200 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **WILKINS, Thea, A.** [US/US]; 262 Palm Avenue, Woodland, CA 95695 (US).

(74) Agents: **HAJOS, Brigitte, A.** et al.; Townsend and Townsend and Crew LLP, Two Embarcadero Center, 8th Floor, San Francisco, CA 94111 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

(88) Date of publication of the international search report: 7 April 2005

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: BIOENGINEERING COTTON FIBER PROPERTIES

(57) Abstract: The present invention provides plant fiber expansion (FE) genes that encode FE polypeptides, such as phosphoenol pyruvate carboxylase (PEPcase), expansin, endoglucanase, xyloglucan endoglycosyltransferase (XET), and pectin methyl esterase (PME). The invention further provides fiber-specific promoters. Still further, the invention provides molecular strategies for modulating fiber quality and yield in fiber producing plants by modulating expression of FE genes or mutant forms of FE genes.

WO 2003/098186 A3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/15269

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C07H 21/04; C12N 15/82, 15/87; A01H 5/00
US CL : 536/23.1, 23.6, 24.1; 800/298, 290, 278, 287; 435/468

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
U.S. : 536/23.1, 23.6, 24.1; 800/298, 290, 278, 287; 435/468

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
Please See Continuation Sheet

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	NCBI database for nucleotide sequences, National Center for Biotechnology Information, National Library of Medicine, NIH (Bethesda, MD, USA) VODJANI, F., et al., Accession Number AF008939, 1997. 100% identical to SEQ ID NO:1.	1-2
AP	US 6,472,588 B1 (HALGLER et al) 29 OCTOBER 2002 (29.10.2002), entire document.	5-8, 10-12, 14-16, 18-21, 24-29
AE	US PUBLICATION NUMBER US 2004/0197915 A1 (MATSUOKA et al) 7 OCTOBER 2004 (7.10.2004), entire document	5-8, 10-12, 14-16, 18-21, 24-29
A	ORFORD, S.J. et al. Specific expression of an expansin gene during elongation of cotton fibres. Biochimica et Biophysica Acta 1398. 1998, Vol. 3, pages 342-346, entire document.	1-8, 10-12, 14-21, 23-29
A	US 5,495,070 (JOHN) 27 FEBRUARY 1996 (27.2.1996), entire document.	1-8, 10-12, 14-21, 23-29

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search
27 October 2004 (27.10.2004)

Date of mailing of the international search report
28 JAN 2005

Name and mailing address of the ISA/US
Mail Stop PCT, Attn: ISA/US
Commissioner for Patents
P.O. Box 1450
Alexandria, Virginia 22313-1450
Facsimile No. (703) 305-3230

Authorized officer
Stuart F. Baum

Telephone No. 703-308-0196

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US03/15269

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claim Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claim Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claim Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: Please See Continuation Sheet

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

PCT/US03/15269

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-2, 5-8, 10-12, 14-16, 18-21 and 24-29, drawn to an isolated nucleic acid molecule comprising a FE polynucleotide of SEQ ID NO:1 encoding SEQ ID NO:2, transgenic plant and method of modulating fiber quality and yield comprising said polynucleotide.

Group II, claim(s) 1-8, 10-12, 14-21 and 23-29, drawn to an isolated nucleic acid molecule comprising a FE polynucleotide of SEQ ID NO:3 encoding SEQ ID NO:4, transgenic plant and method of modulating fiber quality and yield comprising said polynucleotide.

Group III, claim(s) 1-2, 5-8, 10-12, 14-16, 18-21 and 24-29, drawn to an isolated nucleic acid molecule comprising a FE polynucleotide of SEQ ID NO:5 encoding SEQ ID NO:6, transgenic plant and method of modulating fiber quality and yield comprising said polynucleotide.

Group IV, claim(s) 1-2, 5-8, 10-12, 14-16, 18-21 and 24-29, drawn to an isolated nucleic acid molecule comprising a FE polynucleotide of SEQ ID NO:7 encoding SEQ ID NO:8, transgenic plant and method of modulating fiber quality and yield comprising said polynucleotide.

Group V, claim(s) 1-2, 5-8, 10-12, 14-16, 18-21 and 24-29, drawn to an isolated nucleic acid molecule comprising a FE polynucleotide of SEQ ID NO:9 encoding SEQ ID NO:10, transgenic plant and method of modulating fiber quality and yield comprising said polynucleotide.

Group VI, claim(s) 9, 13 and 22, drawn to an isolated nucleic acid molecule of SEQ ID NO:1 operably linked in antisense orientation to a promoter, transgenic plant and method of modulating fiber quality and yield in a plant comprising said nucleic acid molecule in antisense orientation.

Group VII, claim(s) 9, 13 and 22, drawn to an isolated nucleic acid molecule of SEQ ID NO:3 operably linked in antisense orientation to a promoter, transgenic plant and method of modulating fiber quality and yield in a plant comprising said nucleic acid molecule in antisense orientation.

Group VIII, claim(s) 9, 13 and 22, drawn to an isolated nucleic acid molecule of SEQ ID NO:5 operably linked in antisense orientation to a promoter, transgenic plant and method of modulating fiber quality and yield in a plant comprising said nucleic acid molecule in antisense orientation.

Group IX, claim(s) 9, 13 and 22, drawn to an isolated nucleic acid molecule of SEQ ID NO:7 operably linked in antisense orientation to a promoter, transgenic plant and method of modulating fiber quality and yield in a plant comprising said nucleic acid molecule in antisense orientation.

Group X, claim(s) 9, 13 and 22, drawn to an isolated nucleic acid molecule of SEQ ID NO:9 operably linked in antisense orientation to a promoter, transgenic plant and method of modulating fiber quality and yield in a plant comprising said nucleic acid molecule in antisense orientation.

The inventions listed as Groups I-X do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Applicant's chemical compounds, i.e., different DNA sequences encoding different polypeptides, each have different properties and different core structures that elicit different activities, and as such, there is no special technical feature that is linked or is shared between the different nucleic acid or protein sequences. In addition, there is no special technical feature that links or is shared between the nucleic acid sequences encoding polypeptide sequences of Groups I-V and the nucleic acid molecules in antisense orientation of Groups VI-X.

INTERNATIONAL SEARCH REPORT

PCT/US03/15269

Continuation of Box II Item 3:
1-8, 10-12, 14-21, 23-29 including SEQ ID NO:1 encoding SEQ ID NO:2 and SEQ ID NO: 3 encoding SEQ ID NO:4

Continuation of B. FIELDS SEARCHED Item 3:
sequence search for SEQ ID NO: 1 and 3, and DNA encoding SEQ ID NO:2 and 4; WEST, STN; agricola, biosis, caplus, caba